The Importance of Service and Genre in Recommendations for Online Radio and Television Programmes *

Ian Knopke†
British Broadcasting Corporation
201 Wood Lane, White City
London, UK
ian.knopke@gmail.com

ABSTRACT

The BBC iPlayer is an online delivery system for both radio and television content [1]. One of the unique features of the iPlayer is that programming is based around a seven day "catch-up" window. This paper documents some early investigations into features that may be used to produce quality recommendations for that system. The two features explored here, *services* and *genre*, are partly unique to BBC metadata, and are available for all programmes in the schedule. Services are roughly equivalent to channels or stations, while genres are editorially-assigned categorisations of media content. Results of genre / service-based diversity are presented, as well as some simple recommenders based on there, and additional discussion of the topic and results.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Recommendations

Keywords

Recommendations, Broadcasting, Collaborative Filtering

1. INTRODUCTION

The BBC iPlayer is an online delivery system for both radio and television content. Freely available for users within the geographical borders of the United Kingdom, it has been immensely successful and is used by millions of people each day. Unlike similar systems from commercial broadcasters, the BBC's system is provided without advertising.

*(Produces the permission block, and copyright CC 3.0 information). For use with SIG-ALTERNATE.CLS. Supported by ACM.
†

WOMRAD 2011 2nd Workshop on Music Recommendation and Discovery, colocated with ACM RecSys 2011 (Chicago, US)

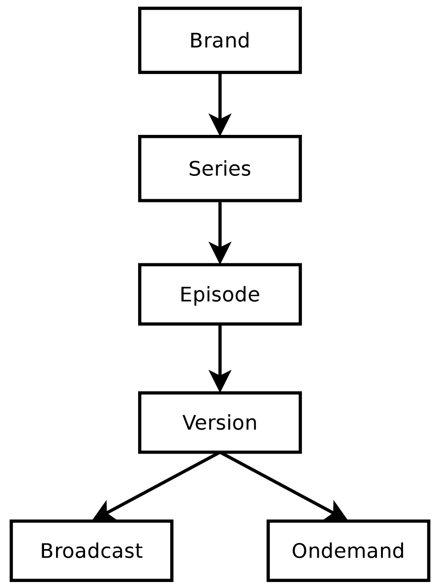

Figure 1: BBC Programme Hierarchy

One of the unique features of the iPlayer is that programming is based around a seven day "catch-up" window. Programming is first shown as a linear broadcast over normal transmission systems (radio and tv). Shortly thereafter the same content becomes available online, without charge, for a period of one week. The BBC maintains a near-perfect synchronicity between their linear and online broadcasting worlds, with over 95% of linear content available as "catch-up" internet television or radio on many different gaming consoles, integrated television platforms and mobile devices, as well as desktop and laptop computers. This synchronicity is completely integrated at both the metadata and transcoding levels, and across both radio and television.

A simplified diagram of the BBC programme metadata hierarchy is shown in Figure 1. The most important element for purposes of this paper is the episode. Episodes may be edited into different versions, and then sent out as transmitted broadcasts or made available online as ondemands. Episodes are grouped into series, under a particular brand. Brands are equivalent to what a user might find listed in a programme guide; common UK examples are EastEnders or Dr. Who (tv) or Desert Island Discs (radio).

This paper documents an investigation into features that may be used to produce quality recommendations. The two features explored here, services and genre, are partly unique to BBC metadata, but genres are also used in other music and media recommendation systems. While there is a large body of research into content-based features for music recommendation, it should be noted that the research presented here is entirely based on metadata, user histories, and the BBC programme hierarchy.

2. RECOMMENDATION SYSTEMS FOR THE BBC IPLAYER

2.1 Previous Issues and Possible Solutions

Most recommendation systems generate recommendations by identifying similar users based on their recorded product choices, and then identifying products popular with these users that a new, similar user has not yet chosen. This is often referred to as *collaborative filtering*. Amazon and last.fm are two examples of such systems [4], and there are many variants [7, 5].

These systems have proven to be effective in many commercial environments, leading to increased site traffic, sales, and an improved connection between individual users and the items that they are interested in. However, a system of this type was recently incorporated into the BBC iPlayer product and failed to produce similar behaviour, with a daily usage rate of approximately 4% of episode click-throughs. It is useful to examine some of the reasons why a technique that has been successful in other online contexts would perform so poorly in the case of the BBC. Particular issues with standard collaborative filtering systems include:

Dynamic Programme Schedule Most online stores have a collection of items, such as books or songs, that are largely unchanging. While new items are often added, the amount of new material in relation to the majority of the collection is small enough that one can consider it to be relatively static. In practice, the relatively small number of new items added can be handled through weekly or daily recalculations of recommendations across the entire product set / user histories. In contrast, the list of iPlayer ondemands is primarily limited to a seven day availability window. The composition of programs within this window changes dynamically, with new programmes being added at least every hour, and older ones expiring. The list of valid programmes, and effectively the viewer's history of programmes to recommend against only extends back a week. In effect, a completely new set of programmes is introduced every week, making it difficult to leverage the user's play history towards generating new recommendations.

Cold Start Problem In the classical collaborative filtering model, new items do not get recommended until enough users have discovered them through other means. This is really just another aspect of the sparse data problem, where there is not enough user history to make adequate recommendations [3]. New items are often introduced to users through mechanisms such as promotions, or through partial solutions such as artificially introducing non-personalised defaults based on average user ratings of all products [2]. In contrast, the short availability window of iPlayer programmes effectively meant that existing programmes never left this "build-up" phase of generating enough history with which to make effective recommendations. In most cases new programmes often weren't recommended until they were near the end of their availability windows. It is an extremely unfortunate situation to have the BBC place considerable effort into creating world-class content, and then not recommend it for the majority of that programme's availability, or perhaps not at all.

Eliminating Old Content In a typical collaborative filtering system, removing items requires recalculation of the mathematical relationships between all users and products (or just products to products). This is a computationally-expensive process, and consequently most online stores only remove products from their catalogues infrequently. If necessary, invalid results can be temporarily filtered until such time as a system-wide batch recalculation can be accomplished. In some cases removal of items can cause referential integrity (foreign key) issues, and many collaborative filtering systems apparently do not have mechanisms for removing content at all. This led to many programmes being recommended that were no longer available, and required the implementation of an expensive, secondary real-time filtering system to remove expired recommendations.

2.2 Possible Solutions

One obvious but partial solution to these problems would be to filter the output results to only produce recommendations within the current time window. While this would alleviate the problem of producing expired recommendations, it does not solve other issues such as the cold start problem.

Another approach, and the one explored here, is to instead find more general categorisations for programmes. If all programmes in the current schedule can be assigned to a set of static categories, these can then be used to record user histories against. The experiments in this paper explore the potential of two such features, services and genre, for use in storing cumulative user histories. These have the advantage of being assigned to all radio and television programmes in the BBC linear schedule and are readily available.

2.3 Services

In linear broadcasting, a service is a particular station or channel such as "BBC One" or "6 Music". In the world of online "catch-up" broadcasting services tend to function more as an association of programmes that share some common heritage. The reasons for this are partly historical, but these divisions are also still valid from an audience perspective; the original channel structures were created to fulfill different audience requirements. For instance, "6 Music" tends to focus on very new music, while the "BBC Four" radio audience is more classically oriented. However, one of the advantages of online broadcasting is that audience members have the ability to switch between services more easily than ever before. When removed from the restraints of the linear schedule, one would expect to see users take advantage of this and new listening trends and patterns to be reflected in user play histories.

2.4 Genres

Table 1: Common BBC Services and Genres

Services	Genres
bbc_1xtra	childrens
bbc_6music	religion_and_ethics
bbc_7	entertainment
bbc_london	drama
bbc_radio_five_live	factual
bbc_radio_one	weather
bbc_radio_three	music
bbc_radio_two	sport
bbc_three	news
bbc_world_service	comedy

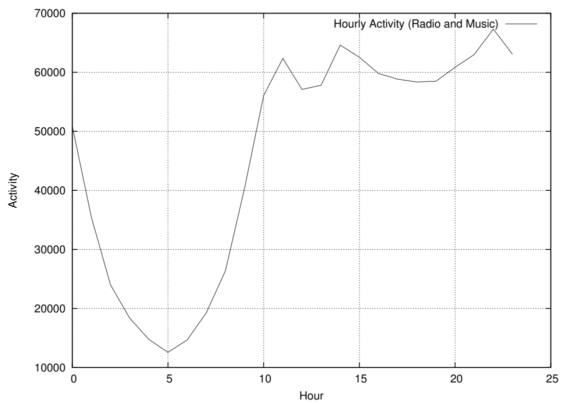

Figure 2: Accumulated Daily Online Radio Activity

Every BBC programme, both television and radio, has at least one genre assigned to it by an expert editorial staff member. These are used in a variety of marketing and promotional functions, as well as for programming, and are considered to be accurate in the broadcasting industry.

A list of some common BBC services and genres is given in Table 1. While the properties of services and genre in relation to the linear broadcast audience is well known, similar information about online usage is not as well evaluated. Both features, however, are thought to be influential in the online domain. The value of these for recommending online programming remains relatively unevaluated in an empirical way.

3. EXPERIMENT

We performed two kinds of experiments. First, the diversity of genres and services were tested. Based on this, four simple recommendation systems were evaluated for how close a match they were to a historical dataset.

A month of iPlayer play history was made available from May 28 to June 25, 2010, consisting of approximately 18 million instances of user selected ondemands, with most shows lasting a half or full hour. Of this, approximately 17 million are televised selections and 1 million are radio. Daily online radio and television usage patterns, averaged over the time period are given in Figures 2 and 3 respectively.

After some discussion and initial exploration, it was decided to test these factors based on the diversity of user play history. To test the diversity of both services and genres, a play history of 89,574 radio and 747,992 television users for

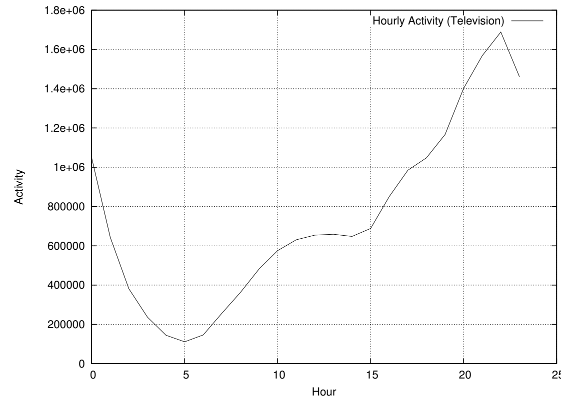

Figure 3: Accumulated Daily Online Television Activity

Table 2: Diversity of BBC Services

	Radio / Music	TV
Gini	0.03	0.25
Entropy	0.07	0.6
Classification Error	0.025	0.19

the above time period were extracted, and the diversity of each user's individual history was calculated. While other more complex diversity evaluation systems are available [6], three common measures of diversity were used: Gini impurity, entropy (2), and a standard classification error using the maximum value (3).

$$\text{gini}(t) = 1 - \sum_{i=0}^{c-1} p(i|t)^2 \tag{1}$$

$$\text{entropy}(t) = -\sum_{i=0}^{c-1} p(i|t) log_2 p(i|t) \tag{2}$$

$$\text{maxclasserror}(t) = 1 - max_i p(i|t) \tag{3}$$

Table 2 shows the averaged values for all users. For comparison purposes, similar figures were also calculated for the television users. These results clearly show that the majority of individual radio users concentrate around a very small number of services, with very little diversity. Television users, on the other hand, tend to have much more diverse service histories and do not appear to be as tied to particular services in the online world. Similar figures for genre are show in Table 3 and to a lesser degree exhibit the same trends.

Based on these results, four simple recommendation strategies were tested for recommending radio and music programmes. Recommendations were based on:

Table 3: Diversity of BBC Genres

	Radio / Music	TV
Gini	0.15	0.37
Entropy	0.32	0.93
Classification Error	0.14	0.31

Figure 4: Markov Chain built from genres using BBC 3

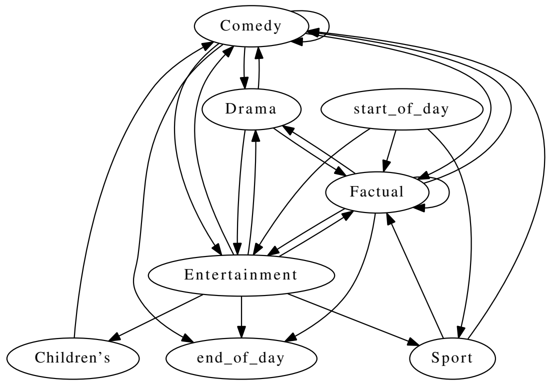

Table 4: Results of Simple Recommenders

Last programme	.06
Most common	.14
Markov	.28
Markov w/services	.34

- The genre of last programme
- The most common genre in the user's history
- A Markov chain of genres derived from all linear broadcast schedules
- Individual Markov chains of genres for each service

The inclusion of Markov chains requires some explanation. The order of programmes is traditionally an important factor in the scheduling of linear broadcasts, with the intention of sustaining audience interest for longer time periods. Consequently, a simple Markov chain based on successive genres was constructed using the linear schedules. Effectively this reduces to a probability distribution for each genre where the most likely genre was compared to that of the next item in the user's history. Note that start and end-of-day states were inserted to represent the 6 AM daily schedule changeover, as no connection is implied between days. In the case of the fourth recommender, individual Markov chains were built for each service and resolved using the service of the previous programme. As an example, Figure 4 shows a simple Markov chain built on successive genres for BBC 3.

Each recommender was then tested on each user's play histories in sequence and a tally of matches / failures kept. These were evaluated using the user's past histories as simple percentages, as shown in Table 4.

4. DISCUSSION

While none of the strategies tested could be considered a complete recommendation system, it is surprising that correct results can be obtained more than a third of the time using only these two simple features, and a knowledge of the programmes found in the linear schedule. One possible way to interpret this is that the online audience shares some of the behaviour of the linear scheduling audience, even when freed of the constraints of only having a single content choice at any one time.

Also, the use of the original service has more of an impact in a radio context than in a television context. To be sure, there are significant differences between television and radio as media formats, and in many ways are not comparable. Nevertheless, it is interesting to try. One possible interpretion is that television viewers have embraced the online experience to a greater extent than pure music or radio listeners. However, it may also be that radio users are more loyal in general to particular stations/brands than television users for other reasons besides just the music. For instance, online radio stations such as last.fm specialise in automatically generating curated collections of music. Disregarding any differences between their recommendations and those programmed by the human curators at the BBC, the main difference is the other elements such as presenters and news segments, and these may be what keeps listeners from changing services.

Genre is also useable for radio recommendations, but genre as a single feature appears to work better for recommending television programmes.

5. FUTURE DIRECTIONS

While this was more on the order of an initial exploration of the problem space, the work presented here suggests a number of additional areas of research. It seems clear that time of day is also probably an important factor. We would also like to do better comparisons between the linear and online audience behaviours, as it seems that there is probably a fair amount of common behaviour there. Also, the study should be expanded to include additional features.

6. REFERENCES

[1] BBC. iPlayer, 2011. `http://www.bbc.co.uk/iplayer/`.

[2] J. S. Breese, D. Heckerman, and C. M. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52, Madison, WI, 1998. Morgan Kauffman.

[3] C.-N. Hsu, H.-H. Chung, and H.-S. Huang. Mining skewed and sparse transaction data for personalized shopping recommendation. *Machine Learning*, 57(1-2):35–59, 2004.

[4] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.

[5] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 285–95, 2001.

[6] M. Slaney and W. White. Measuring playlist diversity for recommendation systems. In *Proceedings of the ACM Workshop on Audio and Music Computing for Multimedia*, pages 22–32, Santa Barbara, CA, USA, 2006. ACM.

[7] X. Su and T. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:1–20, 2009.